Serie Desarrollo del Pensamiento Matemático No. 17

Introducción a la Estadística







Serie Desarrollo del Pensamiento Matemático No. 17

Introducción a la Estadística







por Martín Andonegui Zabala

372.7
And.
Cuaderno N° 17
Introducción a la Estadística
Federación Internacional Fe y Alegría,
junio 2007
32 p.; 21,5 x 19 cm.
ISBN: 978-980-6418-94-3
Matemáticas, Estadística

“La base educativa que debemos consolidar, pensando siempre en el influjo político que alcanzarán un día nuestros alumnos, debe ser de plenitud humana, de fortalecimiento de su personal identidad, de dotación instrumental de los recursos culturales, tecnológicos y creativos que hacen a un hombre más fuerte, más digno y más generoso.”

Padre José María Vélaz

EQUIPO EDITORIAL
Beatriz Borjas y Carlos Guédez

Dimensión: Desarrollo del pensamiento matemático
Cuaderno N° 17
Introducción a la Estadística
Autor: Martín Andonegui Zabala

Este libro se ha elaborado con el propósito de apoyar la práctica educativa de los cientos de educadores de Fe y Alegría. Su publicación se realizó en el marco del ***Programa Internacional de Formación de Educadores Populares*** *desarrollado por la Federación Internacional Fe y Alegría desde el año 2001.*

Diseño y Diagramación: ***Moira Olivar***
Ilustraciones: ***Corina Álvarez***
Concepto gráfico: ***Juan Bravo***
Corrección de textos: ***Carlos Guédez y Martín Andonegui***

Edita y distribuye: Federación Internacional de Fe y Alegría. Esquina de Luneta. Edif. Centro Valores, piso 7 Altagracia, Caracas 1010-A, Venezuela. Teléfonos: (58) (212)5631776 / 5632048 / 5647423.
Fax: (58) (212) 5645096
www.feyalegria.org

© Federación Internacional Fe y Alegría

Depósito legal: lf 60320073102352
Caracas, junio 2007

Publicación realizada con el apoyo de:
Centro Magis - Instituto Internacional para la Educación Superior en América Latina y el Caribe (IESALC) - Corporación Andina de Fomento (CAF)



introducción

A modo de introducción..., nuestro recordatorio

La sugerencia que proponíamos en el Cuaderno No 1 y que siempre presidirá los demás Cuadernos: Vamos a estudiar matemática, pero no lo vamos a hacer como si fuéramos simplemente unos alumnos que posteriormente van a ser evaluados, y ya. No. Nosotros somos docentes –docentes de matemática en su momento- y este rasgo debe caracterizar la forma de construir nuestro pensamiento matemático. ¿Qué significa esto?

• La presencia constante de la meta última de nuestro estudio: alcanzar unos niveles de conocimiento tecnológico y reflexivo, lo cual debe abrir ese estudio hacia la búsqueda de aplicaciones de lo aprendido, hacia el análisis de los sistemas que dan forma a nuestra vida y utilizan ese conocimiento matemático, y hacia criterios sociales y éticos para juzgarlos.

• Construir el conocer de cada tópico matemático pensando en cómo lo enseñamos en el aula, además de reflexionar acerca de cómo nuestro conocer limita y condiciona nuestro trabajo docente. De esta forma, integrar nuestra práctica docente en nuestro estudio.

• Como complemento a lo anterior, construir el conocer de cada tópico matemático pensando en cómo lo podemos llevar al aula. Para ello, tomar conciencia del proceso que seguimos para su construcción, paso a paso, así como de los elementos –cognitivos, actitudinales, emocionales...- que se presenten en dicho proceso. Porque a partir de esta experiencia reflexiva como estudiantes, podremos entender y evaluar mejor el desempeño de nuestros alumnos –a su nivel- ante los mismos temas.

• En definitiva, entender que la matemática es la base de su didáctica: la forma en que se construye el conocimiento matemático es una fuente imprescindible a la hora de planificar y desarrollar su enseñanza.

Y ahora, vamos al tema de este Cuaderno, la Estadística.

1. El significado de la Estadística

Vivimos bombardeados de ***información*** por todas partes; a veces, hasta en exceso. Los medios de comunicación, los libros, lo que nos dicen los demás, lo que observamos en nuestro caminar por la ciudad o el pueblo, lo que recordamos, lo que leemos o estudiamos... Hay información permanente, abundante; quizás, hasta excesiva...

Los elementos que componen esta información son de naturaleza muy diversa: noticias puntuales, ideas, descripciones de objetos y acontecimientos, comentarios, apreciaciones, rumores... También hay otros que se refieren a ciertas ***características*** de personas u objetos, características que son ***susceptibles de clasificación***; por ejemplo, el género (sexo) de las personas, su color de piel, su nacionalidad, su lugar de residencia, sus creencias religiosas, sus inclinaciones políticas...; o las clases de animales, de plantas..., los tipos de viviendas...

Finalmente, hay otros elementos que nos proporcionan información y que son ***susceptibles de algún tipo de medición cuantitativa***, es decir, que pueden presentarse como números que se refieren a múltiples aspectos o fenómenos: costos de los alimentos, de los servicios, de las medicinas...; temperaturas ambientes a lo largo de un día o de un año, o niveles de humedad en las distintas regiones del país; calificaciones de nuestros alumnos; distribución de la población nacional de acuerdo a su género, a sus edades, o a su ubicación por departamentos, provincias o estados; porcentajes anuales de mortalidad infantil; resultados de encuestas de opinión sobre innumerables temas...

Las características de los objetos o personas que son susceptibles de clasificación o de medición cuantitativa suelen presentar rasgos de ***variabilidad***; por ejemplo, existen diversos colores de la piel o de los ojos, diversas nacionalidades, diversos precios de las cosas y de los servicios, diversas calificaciones escolares, diversas estaturas y pesos en las personas, etc. En razón de esta variabilidad, tales características se denominan ***variables***.

Por lo que acabamos de decir, ya vemos que las variables pueden ser de ***dos tipos***:

- ***Cualitativas*** (también llamadas ***atributos***): sólo susceptibles de ***clasificación***.
- ***Cuantitativas***: susceptibles de ***medición numérica***.

Entre estas últimas encontramos dos maneras de medir. La primera se refiere y aplica a las características que se miden por medio del ***conteo***. Por ejemplo, las inasistencias diarias de los alumnos de nuestra escuela; o el número de personas que habitan en la vivienda de cada uno de nuestros alumnos. Este tipo de variables que se miden contando recibe el nombre de ***discretas***. Como se ve, estas variables sólo pueden tomar ***valores enteros***.

La segunda manera de medir se refiere y aplica a las variables que se denominan ***continuas*** porque, entre dos valores fijos, pueden tomar ***todos los valores intermedios*** (decimales). Tal es el caso, por ejemplo, del peso o la estatura de las personas, o la velocidad de un carro...

Claro que, para estas variables, los instrumentos de medición tienen sus limitaciones y siempre se termina por fijar la precisión de las medidas hasta cierto punto. Por ejemplo, en la estatura se llega hasta los centímetros: 1 m y 52 cm, ó 152 cm; y en el peso, hasta los hectogramos: 37,5 Kg; rara vez se detallan con más precisión...

1. Identifique cada una de las siguientes variables como cualitativa o cuantitativa; en este último caso, determine si es discreta o continua:

a) Estado civil
b) Ingresos mensuales familiares
c) Número del documento de identidad
d) Peso de una persona
e) Número de páginas de un libro
f) Profesión u ocupación
g) Estatura de una persona
h) Simpatía por un movimiento político
i) Temperatura ambiente
j) Precio del transporte público

k) Número de personas del grupo familiar
l) Número de calzado
m) Mascota o animal preferido

Conviene diferenciar a las variables cualitativas y cuantitativas por otro rasgo. Las primeras manifiestan su variabilidad en el abanico de categorías en que se clasifican, categorías que no mantienen una ***relación de orden*** entre sí.

Por ejemplo, no es posible establecer cuál de los colores está por encima de los otros, o cómo se ordenan las diversas nacionalidades u ocupaciones... (claro que hay quien lo hace, pero por razones externas a las propias variables: gustos, razones subjetivas, ansias de poder y dominio...).

En cambio, entre los valores que pueden tomar las variables cuantitativas sí existe un cierto orden. Por ejemplo, es posible ordenar los valores de las estaturas, de las edades de las personas, de los números de páginas de los libros, etc.

Por esta razón se dice que las variables cualitativas se miden en una ***escala nominal*** (sus valores se reducen a los "nombres" de las categorías en que se clasifica el atributo) y que las variables cuantitativas se miden en una ***escala ordinal*** (tiene sentido ordenar sus valores).

Dentro de toda esta diversidad, hay algunos rasgos comunes a destacar. En primer lugar, las variables y los resultados de las mediciones se refieren a ***poblaciones***, que son los conjuntos de personas, de objetos, de características ambientales..., que portan alguna información relativa a la variable que se estudia: el género, la edad, la ubicación geográfica, o las opiniones de las ***personas***; la tasa de mortalidad de los ***niños***; los precios de los ***objetos*** y de los ***servicios***; la temperatura o grado de humedad de nuestro ***ambiente***; las calificaciones de nuestros ***alumnos***... Cada uno de los elementos de una población (sean personas, objetos, etc.) recibe el nombre de ***individuo***.

Los resultados de la medición de una variable en el seno de una población reciben el nombre de ***datos.*** Los datos –sean categorías o clases de un atributo, o números- reflejan la variabilidad de la característica estudiada en esa población. En este sentido, los datos son ***mediciones en un contexto específico***, condición indispensable para que puedan transmitir información (Moore, 1998).

Por ejemplo, el número 3,2 no nos aporta información alguna; pero si decimos que una niña pesó 3,2 Kg al nacer, sí podemos establecer algunas interpretaciones acerca de su estado saludable en el momento de su nacimiento. Y si disponemos de una colección de datos similares como, por ejemplo, los pesos de los niños nacidos en la población en estudio durante un lapso de tiempo determinado, tendremos nuevos elementos informativos para interpretar el peso de nuestra recién nacida; por ejemplo, si está dentro de lo normal, o no.

Todos sabemos el nombre de estas listas o colecciones de datos referidos a una variable de una determinada población: son las ***estadísticas propias de esa variable poblacional***. Esta expresión se deriva del término ***Estadística***, utilizado para designar a la ciencia que se ocupa del tratamiento de la información; es decir, de estudiar los fenómenos de cualquier tipo por medio de datos observados y cuantificados, que son recogidos, organizados, representados y analizados con el fin de precisar su significado e inferir, en lo posible, predicciones de cara al futuro. Para todas estas tareas contamos, pues, con ***métodos estadísticos***.

El nombre de ***Estadística*** se atribuye al economista alemán Gottfried Achenwall (siglo XVIII), quien lo derivó del término "estadistas" aplicado a los empleados del Estado dedicados a elaborar los censos poblacionales, (censos de los que se tienen registros históricos mucho más antiguos). Los campos de aplicación iniciales fueron los demográficos y los económicos.

Pero volvamos al hilo conductor de nuestra reflexión. Tenemos que aprender a convivir con la información, a seleccionarla, a quedarnos con la que nos puede ser útil. Porque esa lluvia de información debe tener un objetivo para nosotros; en otras palabras, debemos estar en condiciones de responder a la pregunta: ¿Para qué nos sirve la información?

Probablemente ya tenemos alguna respuesta a esa cuestión. Primero, para mantenernos en conexión con nuestro mundo y con quienes lo comparten con nosotros, familiares, colegas, alumnos, personas de nuestra comunidad, compatriotas... Necesitamos compartir un saber, un conocer; es

preciso estar enterados, no descolgarnos de los demás ni del mundo.

Pero también –y sobre todo- ***necesitamos la información para tomar decisiones***. Y esto, en todos los niveles de nuestra vida: familiar, profesional, social, política, cultural... Si no disponemos de la información adecuada, podemos errar en nuestras decisiones.

De modo que manejar oportunamente la información se ha convertido en una ***competencia imprescindible*** en nuestra vida. Para alcanzar esa competencia, para saber procesar la información, es muy conveniente que ésta venga expresada de una manera organizada. Así se nos facilitará la tarea de interpretarla como es debido y tomar las decisiones del caso.

Pero no siempre la información circula del entorno hacia nosotros. También camina en sentido inverso, también nosotros somos productores de información. Y en este caso, sigue siendo válida la necesidad de expresarla de manera organizada, para que sus receptores puedan interpretarla sin dificultad y sentirse apoyados en su toma de decisiones.

De las consideraciones anteriores podemos derivar las razones del interés que la Estadística puede tener para nosotros, no sólo como ciudadanos sino, particularmente, como docentes. A este respecto, recordemos todos los registros que anualmente efectuamos en la escuela, referidos a diversas características de nuestros alumnos (grado, género, edad, aspectos referidos a su desarrollo físico y a su salud, características familiares, condiciones socioeconómicas, anotaciones acerca del rendimiento escolar...).

Si nos limitásemos al simple registro y posterior archivo de esos datos, la información recolectada no tendría ninguna relevancia ni trascendencia. Pero si, por el contrario, organizamos y analizamos las diversas relaciones entre ellos, podremos establecer conclusiones interesantes en relación con los grupos estudiados, así como compararlos entre sí. E, incluso, podremos llegar a ver que las regularidades que presentan algunas de esas características nos permitirán inferir conclusiones acerca de los rasgos que tendrán las poblaciones de estudiantes futuros.

Por otro lado, resulta muy importante la tarea de lograr que nuestros alumnos comprendan y sepan utilizar las herramientas estadísticas, con el fin de facilitarles la interpretación de la información que les llega, la organización y presentación de la que ellos pueden producir, y la correcta toma de decisiones.

Estamos, pues, a las puertas de un campo bien interesante, bien útil, bien sencillo... y asequible desde los primeros niveles de escolaridad. No pensemos, de entrada, en cosas complicadas. El requisito básico para adentrarnos en la Estadística es tener hábitos de organización. Con esta predisposición no nos debe costar ir adquiriendo las técnicas correspondientes que, por lo demás, son muy lógicas y asequibles.

2. ¿Qué hacemos con los datos?

2.1. En primer lugar, reconocer su necesidad

Quizá la proposición de –antes que nada- reconocer la necesidad de los datos puede llamar la atención de más de un(a) lector(a). Porque estamos acostumbrados a la idea de que las tareas estadísticas significan exclusivamente la recolección de datos, la elaboración de tablas, de gráficos, el cálculo de la media, la determinación de la moda...; es decir, en Estadística hay que “hacer cosas” con los datos y para ello existen una serie de fórmulas y procedimientos a utilizar correctamente.

Sin embargo, si queremos hacer “nuestra” la Estadística, tenemos que empezar por indagar nuestra actitud ante las situaciones de la vida real; tenemos que preguntarnos si deseamos comprenderlas para manejarlas adecuadamente y tomar las decisiones más pertinentes frente a ellas.

Sobre esta base se apoya la Estadística; y, en particular, sobre la hipótesis de que muchas situaciones de la vida real sólo pueden ser comprendidas a partir del análisis de un conjunto de datos de carácter cuantitativo que hayan sido recogidos en forma adecuada.

Así, como decíamos en el ejemplo anterior, el peso de 3,2 kg de una niña al nacer cobra todo su sentido cuando "disponemos de una colección de datos similares como, por ejemplo, los pesos de los niños nacidos en la población en estudio durante un lapso de tiempo determinado". Lo mismo podemos decir de la consideración del peso y la estatura, o de las calificaciones escolares, de uno solo de nuestros alumnos.

Ahora bien, tenemos que reconocer que la observación de un caso aislado, o recurrir a la experiencia de una sola persona, o intentar encontrar evidencias determinantes en un suceso anecdótico, puede confundirnos y llevarnos a una toma de decisiones equivocada (Batanero, 2002).

Por ello, resulta muy importante percibir y valorar la variación que afecta a esas características (variables) que se manifiestan en una forma cuantitativa en nuestro entorno. Los datos nos hablan de la presencia de la variabilidad en muchas de las situaciones que conforman nuestra vida. De entrada, percibimos que no todo es uniforme, hecho con una sola medida (afortunadamente).

Ante este panorama, la Estadística nos ofrece algunos métodos para intentar comprender y analizar, en lo posible, esta variación, aunque sin negar ni agotar su riqueza y su originalidad. Dentro de ciertos límites, los métodos estadísticos nos permiten buscar explicaciones y causas de la variación y, sobre todo, aprender del contexto.

Reúnase con sus compañeros(as) y trate de hacer una lista de aquellas variables de su entorno cuyos datos pueden resultar de interés para su desempeño profesional.

2.2. En segundo lugar, recolectarlos

¿Dónde recolectamos los datos? Bueno, hay muchos datos que se producen regularmente, con una periodicidad al menos anual. En cada uno de nuestros países, todos los ministerios y despachos gubernamentales elaboran sus Anuarios, con los datos y tablas referentes a sus actividades y a las variables poblacionales que son de su incumbencia. En todos ellos existen también instituciones nacionales de Estadística, encargadas de recopilar los datos estadísticos de interés para estudiosos y curiosos... En el campo educativo, desde cada escuela y desde cada instancia municipal, regional y nacional, se producen cantidades de datos propios de este ámbito educativo.

Incluso, son muchos los bancos de datos que son publicados por organismos internacionales, en los cuales se recogen y comparan los que se producen en cada uno de los países involucrados. En lo que a Latinoamérica se refiere, tenemos a la mano los bancos de datos presentados por la UNESCO, la OEA, la CEPAL-ECLAC (2005), etc., entre otros muchos. La lectura de estos cuadros estadísticos nos permite, no sólo asimilar la información que contienen sino, además, observar la forma en que se presentan.

Pero, como hemos dicho antes, nosotros mismos podemos señalar alguna variable que nos interesa en nuestro desempeño profesional y, seguramente, también podemos recolectar datos referentes a esa variable. Tomemos, por ejemplo, el número de inasistencias diarias de alumnos a nuestra escuela durante un mes (20 días). Estos datos pueden proporcionarnos cierta información acerca de la vida escolar, sobre todo si se comparan con los de otros meses del año escolar; es decir, tienen interés para nosotros.

Supongamos que estos son los datos, día a día, empezando por el primer lunes: {15, 19, 18, 18, 17, 17, 11, 13, 19, 18, 20, 21, 23, 26, 24, 21, 20, 17, 15, 12}.

2.3. En tercer lugar, organizar su presentación

¿Qué hacemos con esos 20 datos (N = 20)? Podemos hacer muchas cosas; por ejemplo:

1**. **Ordenarlos de menor a mayor: {11, 12, 13, 15, 15, 17, 17, 17, 18, 18, 18, 19, 19, 20, 20, 21, 21, 23, 24, 25}. O, también, de mayor a menor. Eso nos permite observar y anotar los ***valores mayor y menor***: 26 y 11, respectivamente. O el valor (los dos valores, en este caso) que se encuentra(n) ***en el medio*** de los datos. O el (los) que ***más veces se repite(n)***...

2. Presentarlos en una tabla en la que se señale cada número de inasistencias (***dato***) con el número de veces (***frecuencia***) con el que aparece (***distribución de frecuencias***):

Datos	Frecuencias (N = 20)
11	1
12	1
13	1
15	2
17	3
18	3
19	2
20	2
21	2
23	1
24	1
26	1

3. Como en la tabla anterior aparecen muchos datos diferentes (12), podemos pensar en agruparlos un poco y presentarlos en otra tabla en la que se señalen las inasistencias agrupadas por ***intervalos o clases***, con el número de veces en que se presentan (***distribución de frecuencias para datos agrupados***):

Datos agrupados	Frecuencias (N = 20)
11 al 14	3
15 al 18	8
19 al 22	6
23 al 26	3

Quizás al (a la) lector(a) se le puedan ocurrir otras formas de representar los datos anteriores, o algunas particularidades de los mismos; pero vamos a quedarnos con las tres señaladas. Veamos algunas características de algunas de ellas.

La primera nos permite determinar el ***rango o recorrido de la distribución***, es decir, determinar la amplitud del intervalo en que se mueven los valores. La forma de calcularlo es restar los valores extremos; en nuestro ejemplo, el rango de los datos es: 26 – 11 = 15.

Ese resultado ***no coincide con el número de valores diferentes*** que podrían aparecer entre los valores extremos de la colección de datos. Para obtener este último número volvemos a restar estos valores extremos... y sumamos 1 unidad; en nuestro ejemplo es: 26 – 11 + 1 = 16. Si alguien duda por qué se debe sumar 1 unidad, póngase a contar cuántos valores seguidos hay desde 11 hasta 26, ambos incluidos.

Esa primera forma de representación ordenada de los datos va a tener utilidad más adelante, a la hora de analizar la distribución de los mismos, mediante la determinación de ciertos valores más representativos de la misma.

La segunda forma de representar los datos (distribución de frecuencias) es la que mejor y con más detalle nos permite resumir todos los datos y volverlos a "leer" organizadamente. La tercera forma (distribución de frecuencias para datos agrupados) es, a su vez, un resumen de la anterior, y es muy útil cuando los datos son muy numerosos y están muy dispersos o presentan frecuencias muy bajas. Se pierde algo del detalle que ofrece la forma de representación anterior, pero nos permite una lectura más sintetizada de la distribución de los datos.

Suponga que en el siguiente mes, que tiene 23 días lectivos, se recolectan estos datos de inasistencia diaria de alumnos a la escuela: {12, 11, 12, 13, 14, 13, 15, 16, 14, 15, 18, 16, 19, 15, 21, 20, 24, 23, 22, 23, 23, 23, 22}.

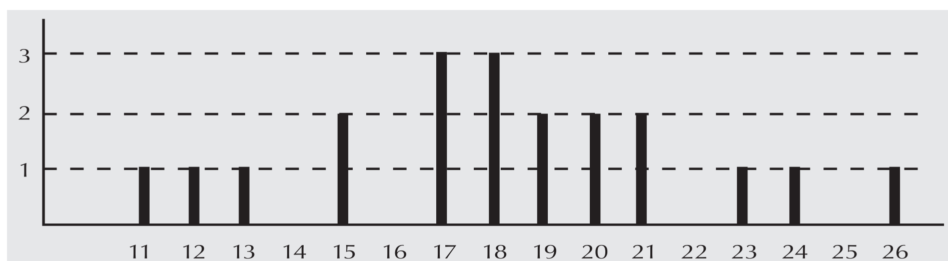
a) Determine los valores extremos de la distribución
b) Calcule el rango de la distribución
c) Ordene los datos de mayor a menor
d) Elabore la tabla de distribución de frecuencias correspondiente
e) Forme intervalos o clases de amplitud 4 (de 10 a 13, etc.) y elabore la correspondiente tabla de distribución de frecuencias

La ***representación*** de los datos, particularmente las referentes a la distribución de frecuencias y a la distribución de frecuencias de datos agrupados, puede hacerse

también de ***forma gráfica*** y no sólo mediante las tablas del tipo que acabamos de presentar (forma tabular).

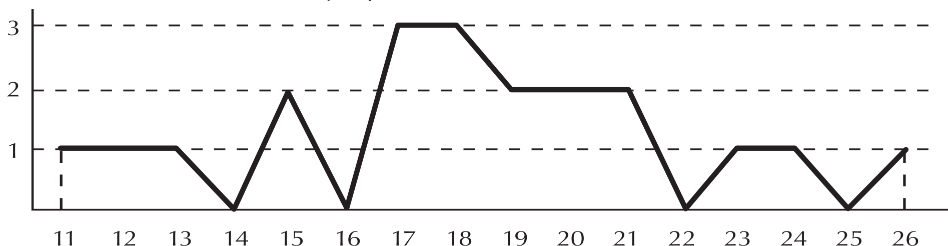
Tomemos el ejemplo de la distribución de frecuencias antes presentada. Para llevarla a una gráfica, vamos a trazar dos segmentos perpendiculares (ejes), con sus marcas o medidas correspondientes; en el horizontal vamos a colocar, ordenados, los valores de la variable (número de inasistencias diarias) y en el segmento vertical, las frecuencias de esos valores.

Marcados así esos dos segmentos, vamos a levantar, en cada punto del eje horizontal, un segmento perpendicular cuya medida sea exactamente la de la frecuencia correspondiente a ese valor de la variable. Haciendo esta tarea para cada uno de los valores de la variable, llegamos a esta gráfica:

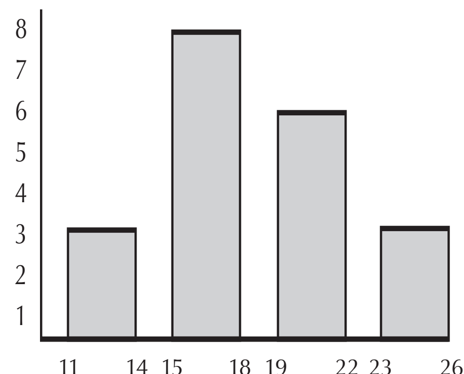


Esta representación recibe el nombre de ***gráfica de barras***. Obsérvese que para conservar la secuencia de los valores de la variable, desde 11 hasta 26, se han marcado los valores 14, 16, 22 y 25, cuya frecuencia es 0. También puede "invertirse" la gráfica, es decir, colocar en el eje vertical los valores de la variable y en el eje horizontal, las frecuencias de esos valores.

Si se toman solamente los vértices superiores de los segmentos o barras verticales correspondientes a las frecuencias, se construye una nueva gráfica, conocida como ***polígono de frecuencias***. Para nuestro ejemplo:



También puede construirse una ***gráfica de barras*** correspondiente a la tabla que recoge la distribución de los ***datos agrupados en clases***. Para nuestro ejemplo:



Tome ahora los datos y las tablas correspondientes a las inasistencias del segundo mes, y elabore:

a) la gráfica de barras
b) el polígono de frecuencias
c) la gráfica de barras para los datos agrupados en clases (10 a 13, etc.)

Hasta ahora hemos trabajado con una variable (inasistencias diarias de los alumnos a la escuela) discreta. Veamos un ejemplo para una variable continua, como el peso o la estatura de las personas.

En nuestro medio escolar, el conocimiento (y seguimiento) del peso y de la estatura de nuestros alumnos tiene su importancia, ya que éstos se encuentran en pleno proceso de desarrollo; vale la pena, pues, recolectar los datos correspondientes e, incluso, hacerlo periódicamente.

En cuanto a la organización de su presentación, que es lo que ahora nos ocupa, enseguida percibimos que lo más práctico es distribuir los datos individuales en clases o intervalos ya que, con seguridad, habrá muchos datos diferentes. De aquí se sigue que las representaciones más apropiadas para estos datos serán la tabla y la gráfica correspondientes a la distribución de datos agrupados.

Supongamos que los pesos de 31 alumnos de una clase son (en Kg): {37,8; 35,6; 34; 31,9; 40,5; 34,2; 35,6; 38,7; 32,8; 35,4; 41,6; 39,8; 34,5; 37; 42; 36,6; 31,9; 36,5; 35,7; 36; 38; 44,1; 37,2; 36,8; 35; 33,5; 38,9; 37,5; 34; 36,5; 42,5}.

La primera actividad de observación nos lleva a buscar los valores extremos; éstos son: 31,9 y 44,1 Kg. Ahora podemos decidir en cuantas clases o intervalos dividimos el grupo, tomando en cuenta que tampoco nos conviene tener un número excesivo de clases; para nuestro caso, un número apropiado puede ser 4 ó 5 clases.

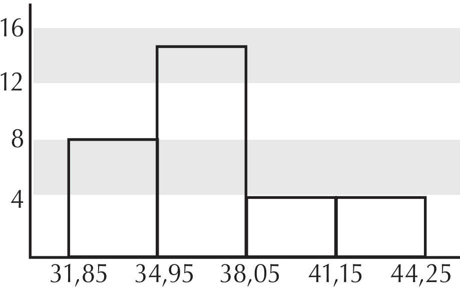
Supongamos que optamos por 4 clases. Esto significa que el gran intervalo entre los valores extremos (de 31,9 a 44,1), cuya diferencia es 12,2, debe repartirse en cuatro lotes de igual tamaño. El tamaño de cada intervalo debe ser, pues, el cociente de 12,2 / 4, que es 3,05, que puede redondearse a 3,1.

Esto significa que si el primer intervalo empieza en el valor menor (31,9), el segundo debe empezar en 31,9 + 3,1 = 35; el tercero debe empezar en 35 + 3,1 = 38,1; y el cuarto, en 38,1 + 3,1 = 41,2. A partir de estos valores deducimos cuáles son los valores mayores de cada intervalo: 34,9; 38; 41,1 y 44,2 (observe que la diferencia entre dos de estos valores seguidos es también 3,1.

Ahora podemos construir la tabla que representa la distribución de datos agrupados:

Datos agrupados	Frecuencias (N = 31)
De 31,9 a 34,9	8
De 35 a 38	15
De 38,1 a 41,1	4
De 41,2 a 44,2	4

y la gráfica correspondiente, que en el caso de las variables continuas, se denomina ***histograma*** (histos [tejido] + gramma [gráfico]):



Obsérvese que en el histograma los rectángulos se adosan unos a otros, ya que la variable es continua. Por esta razón, en el eje horizontal se colocan valores divisorios de las sucesivas clases o intervalos. ¿Cómo se obtiene cada uno de ellos? Se toman dos valores sucesivos, el final de una clase y el inicial de la clase siguiente, y se busca el valor intermedio.

Por ejemplo, 34,95 es el valor intermedio entre 34,9 (valor final del primer intervalo) y 35 (valor inicial del segundo intervalo). Calculado uno de estos valores divisorios, los demás pueden obtenerse agregando o restando el valor de la amplitud del intervalo (3,1). Así se llega, en particular, a los valores divisorios extremos (31,85 y 44,25).

Recolecte los datos de peso y de estatura de los alumnos de su clase y elabore, para cada variable, la tabla y la gráfica de la distribución de datos agrupados correspondiente.

2.4. En cuarto lugar, analizarlos

Los dos puntos anteriores, referidos a la recolección de datos y a la organización de su presentación, no tienen ningún sentido si no desembocan en un análisis de los mismos. Porque recordemos que los datos –y la información que se deriva de ellos- tienen como objetivo facilitar nuestra toma de decisiones.

El análisis de los datos es un campo muy complejo, cuyo desarrollo ocupa el espacio más extenso e importante de la Estadística (lo que se llama la Estadística Inferencial). En este Cuaderno no vamos a entrar en él. Pero sí vamos a proponernos analizar y sacar conclusiones de estas primeras formas sencillas de organizar los datos (Estadística Descriptiva), con el fin de percibir cómo la información ya fluye realmente de ellos.

Volvamos a nuestro primer ejemplo, el de las inasistencias diarias de los alumnos a la escuela, y las tres formas en que organizamos su presentación (ordenados de menor a mayor, la tabla de distribución de frecuencias, y la tabla de distribución de frecuencias para datos agrupados).

La primera forma de organización nos sugiere que el número de inasistencias diarias, durante el mes estudiado, oscila entre una y dos docenas, aproximadamente; que es irregular, es decir, que varía de día en día; y que el número más frecuente de inasistencias ha sido de 17 y 18 (3 veces cada uno).

La tabla de distribución de frecuencias nos permite volver a encontrar y, además, ampliar la información anterior. Por ejemplo, nos permite averiguar en cuántos días del mes se contaron 20 ó más inasistencias (este dato puede ser de interés si, por ejemplo, 20 es un número crítico para la escuela, en el sentido de que cuando se alcanzan o superan las 20 inasistencias diarias, se determina una emergencia escolar...).

¿Cómo calcular ese número de días del mes en que se contaron 20 ó más inasistencias diarias? A partir de la tabla de distribución de frecuencias, basta con sumar las frecuencias correspondientes a los datos 20, 21, 23, 24 y 26 (2 + 2 + 1 + 1 + 1 = 7): durante 7 días las inasistencias llegaron a 20 o estuvieron por encima de ese número.

Como se ve, la tabla de distribución de frecuencias puede servirnos para obtener información nueva referida a la acumulación de frecuencias entre ciertos valores de la variable. Por ejemplo, podemos preguntar cuántos datos estuvieron por debajo de tal valor, o entre dos valores dados, o fueron superiores a determinado valor.

Para responder más directamente a este tipo de tareas, podemos construir una nueva tabla de distribución de frecuencias en la que aparece una nueva columna, la referente a las ***frecuencias acumuladas***; es decir, para cada dato, además de su frecuencia propia, vamos a anotar la suma de las frecuencias correspondientes a los datos inferiores, más la suya propia. En nuestro ejemplo sería:

Datos agrupados	Frecuencias (N = 20)	Frecuencias acumuladas
11	1	1
12	1	2
13	1	3
15	2	5
17	3	8
18	3	11
19	2	13
20	2	15
21	2	17
23	1	18
24	1	19
26	1	20

Como se puede observar, la frecuencia acumulada correspondiente al valor más alto debe coincidir con el total de la población estudiada (20). Si queremos saber durante cuántos días el número de inasistencias diarias estuvo por debajo de 18, basta con leer la frecuencia acumulada correspondiente a 17: durante 8 días. Y para saber durante cuántos días estuvo por encima de 19 (la pregunta inicial), basta restar de 20 la frecuencia acumulada correspondiente a 19: 20 – 13 = 7.

2. ¿Tiene algún sentido elaborar una tabla con frecuencias acumuladas cuando se trata de una variable cualitativa? ¿Por qué?

Siguiendo con el análisis de las inasistencias diarias, nos damos cuenta de que la tabla de distribución de frecuencias para datos agrupados no nos aporta mayor información adicional: solamente que la mayoría de los datos (8 de 20) está en el intervalo de 15 a 18 inasistencias diarias [Notemos que también es posible elaborar la columna de frecuencias acumuladas para el caso en que los datos vengan agrupados en clases].

Hemos utilizado tres de las herramientas habituales de organización de los datos para analizar los datos relativos a una variable (también podíamos habernos servido de las representaciones gráficas correspondientes...). Sin embargo, esto no significa que hayamos agotado todas las posibles vías de su representación y análisis subsecuente.

Por ejemplo, podíamos haber separado los datos de inasistencias en lotes de a cinco días, con el fin de destacar lo que ocurrió cada semana, de lunes a viernes: (15, 19, 18, 18, 17), (17, 11, 13, 19, 18), (20, 21, 23, 26, 24), (21, 20, 17, 15, 12). O también de esta manera:

Semana	Nº de inasistencias				
	Lun	Mar	Mié	Jue	Vie
1ª	15	19	18	18	17
2ª	17	11	13	19	18
3ª	20	21	23	26	24
4ª	21	20	17	15	12

La organización de los datos en esta tabla de doble entrada nos permite ***otro tipo de análisis:***

• Comparar las semanas: Desde el final de la 2ª semana y, particularmente, durante la 3ª, hubo un notable incremento de inasistencias; esto nos dice que la tercera semana ocurrió algo particular, actuó una causa especial (alguna enfermedad, inclemencias climáticas...) cuyo efecto posiblemente fue menguando durante la última semana....

Día/SEMANA	Inasistencias
Lun. 1era sem	15
Lun. 2da sem	17
Lun. 3era sem	20
Lun. 4ta sem	21
Mar. 1era sem	19
Mar. 2da sem	11
Mar. 3era sem	21
Mar. 4ta sem	20
Mie. 1era sem	18
Mie. 2da sem	13
Mie. 3era sem	23
Mie. 4ta sem	17
Jue. 1era sem	18
Jue. 2da sem	19
Jue. 3era sem	26
Jue. 4ta sem	15
Vie. 1era sem	17
Vie. 2da sem	18
Vie. 3era sem	24
Vie. 4ta sem	12

• Comparar los mismos días de cada semana: Por ejemplo, los lunes no han sido, durante este mes, los días críticos para la inasistencia escolar; tampoco los viernes. No se manifiestan hábitos de ausencia alrededor del fin de semana...

Como se puede apreciar, ninguna de estas informaciones nos ha sido proporcionada por las herramientas estadísticas de organización de los datos que hemos manejado con anterioridad. Esta afirmación nos reafirma en el convencimiento de que el último tipo de organización de los datos que hemos manejado es pertinente para su registro y análisis, ya que se trata de una secuencia de datos en el tiempo y, en este caso, es de suma importancia no perder el carácter temporal de la información.

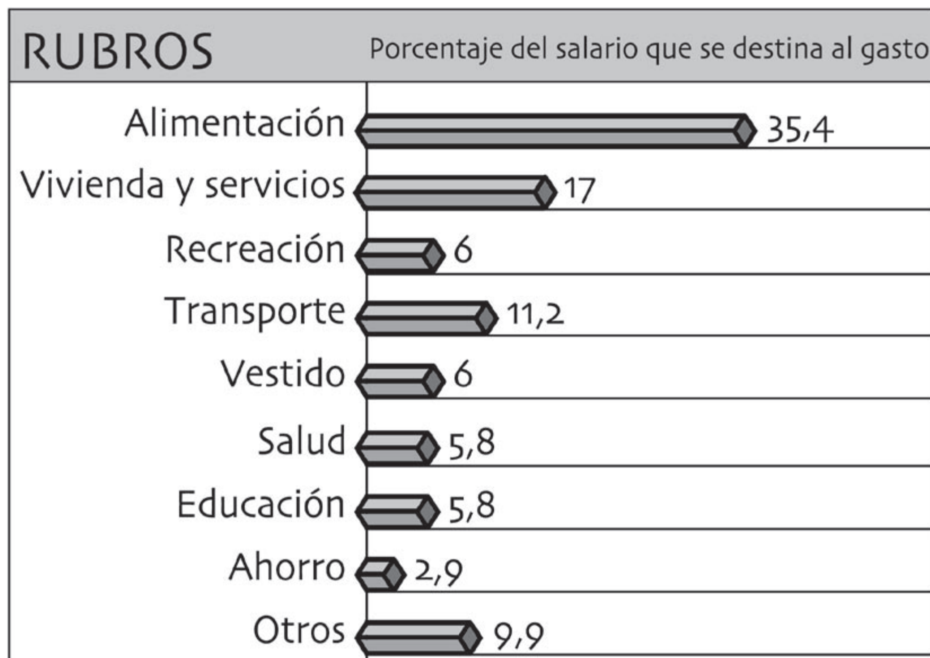
De todo lo anterior queremos extraer algunas conclusiones importantes:

• Hay que saber seleccionar, en cada caso, los instrumentos estadísticos de organización de los datos que nos puedan aportar información pertinente. En nuestro ejemplo, no nos interesa la distribución de frecuencias para los datos agrupados en clases, ya que no aporta nada adicional.

• No debemos cerrarnos a la posibilidad de construir y utilizar otras formas adicionales de organización de los datos, con el fin de extraer información valiosa. En nuestro ejemplo, la tabla de doble entrada resultó ser un instrumento de registro que permitió obtener información adicional muy pertinente.

Tome el ejemplo de los pesos de los 31 alumnos antes considerado, analice los datos presentes en la tabla correspondiente y llegue a algunas conclusiones.

He aquí una gráfica de barras relativa a la *distribución porcentual del ingreso familiar para atender a los gastos correspondientes a los rubros indicados* (datos de uno de nuestros países latinoamericanos, año 2003):



Obsérvese que la variable *distribución porcentual del ingreso familiar es cualitativa, y que sus categorías son, precisamente los rubros o áreas de necesidad indicadas*. Por otro lado, las barras son horizontales y, en lugar de las frecuencias de cada valor de la variable, aparecen los porcentajes correspondientes.

¿Qué información podemos inferir de esta gráfica? Pues, entre otras cosas:

• Un poco más de la tercera parte del ingreso (35,4 %) se dedica a alimentación; este rubro aparece como prioritario.

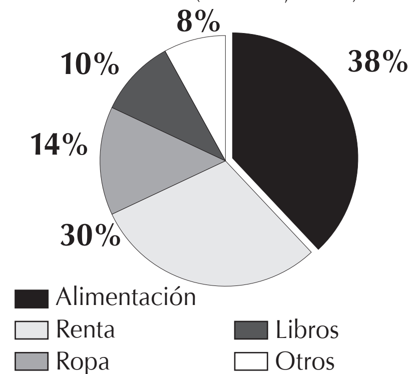
• Los rubros de alimentación, vivienda y servicios, y transporte son los que absorben casi las dos terceras partes (63,6 %) del ingreso familiar.

• El ahorro prácticamente no cuenta para las familias; puede presumirse que no hay cultura de ahorro o que, con más seguridad, las necesidades básicas son tan perentorias que no existe la posibilidad de ahorrar.

• En resumen, se trata de la distribución propia de un país en el que la población destina el ingreso familiar prácticamente para intentar garantizar la supervivencia del día a día...

La información contenida en esta gráfica puede ser todavía mayor si, por ejemplo, los datos se recolectan periódicamente, lo que permitiría un seguimiento temporal con el fin de analizar las variaciones poblacionales en la distribución de su ingreso familiar.

Cuando se trata de una ***variable cualitativa*** cuya distribución en categorías viene cuantificada en porcentajes, puede utilizarse también otro tipo de representación, denominada ***gráfica circular o de pastel.*** Veamos un ejemplo similar al anterior, en el que se muestra la distribución de los gastos mensuales de un estudiante universitario fuera de su casa (Shadian, 1998):



Como se ve, los porcentajes no se representan como barras de diferente longitud, sino como sectores circulares de diferente área (ver Cuaderno 15). Las áreas de estos sectores son proporcionales a los números que indican los porcentajes; a mayor (menor) porcentaje, mayor (menor) área del sector, es decir, mayor (menor) abertura del ángulo central correspondiente.

Para determinar la medida de este ángulo central podemos utilizar una regla de tres muy sencilla:

Porcentaje (%)	Medida del ángulo central del sector circular
100	360^{o}
p	*a*

De esta forma, si conocemos el porcentaje correspondiente a una categoría de la variable, obtendremos el ángulo del sector circular así: $a = \frac{px360}{100}$. En caso de que conozcamos este ángulo, podemos calcular el porcentaje correspondiente de esta forma: $p = \frac{ax100}{360}$.

En el ejemplo, verificamos primero que la suma de los porcentajes sea 100 (hágalo). A partir de ahí, al porcentaje del 38 % le corresponde un ángulo central cuya medida es a = (38 x 360)/100 = 136,8^{o}; y así se calculan todas las demás medidas. Luego, se traza una circunferencia y, con ayuda de un transportador, se van montando los sucesivos ángulos centrales adosando cada uno al anterior.

3. a) Si el estudiante del ejemplo anterior necesita 20.000 pesos en un mes dado, ¿cuánto gasta en alimentación durante ese mes?
b) Si durante otro mes mantiene sus porcentajes de gastos y compra ropa por 2.100 pesos, ¿a cuánto ascienden sus gastos de ese mes en libros?

Hagan una pequeña encuesta en su escuela pidiendo por separado, a los niños y a las niñas, que indiquen cuál es el deporte de su preferencia. En cada uno de los dos casos, tomen los cuatro deportes más destacados y agrupen los demás como “Otros”. Conviertan las frecuencias de selección en porcentajes (la suma de éstos debe ser igual a 100).

Con esta información, construyan las dos gráficas circulares correspondientes. A partir de ellas, analicen las preferencias manifestadas por los alumnos (niños y niñas), por separado y comparativamente; valoren los pros y contras (si existen) de la práctica de cada uno de los deportes señalados (individualismo vs. labor de equipo; posibilidades de inclusión o de exclusión de una porción significativa de alumnos; propensión o no hacia la violencia o hacia la tolerancia; excesiva o moderada competitividad; habilidades que se desarrollan...).

Y saquen sus conclusiones acerca de las posibilidades que tiene la escuela para satisfacer los deseos de los niños y niñas; por ejemplo, ¿existen las instalaciones, los preparadores y los implementos que esos deportes requieren?; ¿se organizan periódicamente competiciones o campeonatos de esas disciplinas deportivas?; ¿se valoran y estimulan los logros de los alumnos y alumnas en sus prácticas deportivas?...¿cómo mejorar las situaciones deficientes?

3. Las medidas de tendencia central

Ya hemos visto lo que podemos hacer con los datos: valorarlos, recolectarlos, organizar su presentación y analizarlos. Vamos a verlos ahora desde otra perspectiva. Por ejemplo, podemos preguntarnos si existe(n) algún(os) dato(s) que sea(n) como especial(es), que pueda(n) mostrarnos alguna característica muy particular de nuestra colección de datos o que pueda(n), incluso, representar a la colección completa.

En el punto 2.3., al presentar los datos ordenados de menor a mayor (o viceversa), decíamos que tal orden nos permite observar y anotar los ***valores mayor y menor***, o el valor que se encuentra ***en el medio*** de los datos, o el (los) ***que más veces se repite(n)***... A estos valores particulares podemos agregar otro, de uso muy frecuente: el ***valor promedio*** de todos los datos.

3.1. La media

El ***valor promedio de todos los datos poblacionales*** recibe el nombre de ***media***. Existen diversos tipos de valores promedio, pero aquí nos vamos a referir a la media aritmética. Para obtenerla, se suman todos los datos y esta suma se divide entre el número de datos considerados.

4. ¿Tiene sentido calcular la media de los datos cuando la variable es cualitativa? ¿Por qué?

En realidad, la media es el valor que cada sujeto de la población tendría si se repartiera "equitativamente" entre todos el valor de la suma total de los datos de la población. La media ***puede coincidir o no con alguno de los datos*** del conjunto.

a) Sea el conjunto de datos: {11, 7, 10, 9, 10, 8, 7, 10}

La media es: (11+7+10+9+8+7+10) / 8 = 72 / 8 = 9

La media coincide con uno de los datos del conjunto

b) Sea el conjunto de datos: {11, 8, 10, 8, 10, 8, 7, 10}

La media es: 72 / 8 = 9

La media no coincide con ninguno de los datos del conjunto

c) Sea el conjunto de datos: {10, 11, 12, 13, 14, 15, 16, 17, 18, 19}

La media es: 145 / 10 = 14,5

La media no coincide con ninguno de los datos del conjunto

5. Calcule la media de inasistencias para los datos del ejemplo inicial: {15, 19, 18, 18, 17, 17, 11, 13, 19, 18, 20, 21, 23, 26, 24, 21, 20, 17, 15, 12}.

6. Calcule la media de los pesos (en Kg) de los niños del ejemplo dado anteriormente: {37,8; 35,6; 34; 31,9; 40,5; 34,2; 35,6; 38,7; 32,8; 35,4; 41,6; 39,8; 34,5; 37; 42; 36,6; 31,9; 36,5; 35,7; 36; 38; 44,1; 37,2; 36,8; 35; 33,5; 38,9; 37,5; 34; 36,5; 42,5}.

Con el fin de presentar una fórmula generalizada para calcular la media de cualquier conjunto de datos, procedemos así:

• Denotamos con n el número de datos del conjunto (31, en el último ejercicio).

• Denotamos los n datos del conjunto con los símbolos $x_1, x_2, \dots, x_n$, donde, por ejemplo, x_3 representa el tercer dato (34 Kg, en el último ejercicio) y x_{18} es el que ocupa la posición 18 (36,5 Kg, en el último ejercicio).

• Si denotamos la media con el símbolo $\bar{x}$, la fórmula que permite calcular su valor se expresa así:

$$\bar{x} = \frac{x_1 + x_2 + x_3 + \dots + x_n}{n} \quad [1]$$

Como se ve, para calcular la media no hace falta que los datos estén ordenados, ya que la suma es conmutativa.

Por otro lado, veamos el ***caso en que los datos se presentan en una tabla de distribución de frecuencias***, como la del primer ejemplo de inasistencias a la escuela:

No. Inasistencias	Frecuencia	Frec. Acumulada
11	1	1
12	1	2
13	1	3
15	2	5
17	3	8
18	3	11
19	2	13
20	2	15
21	2	17
23	1	18
24	1	19
26	1	20

Para calcular la media efectuaríamos la suma: 11 + 12 + 13 + 15 + 15 + 17 + 17 + 17 + ...; está claro que podemos abreviar la suma si los sumandos que se repiten se colocan en forma de producto; por ejemplo: 11 + 12 + 13 + (2 x 15) + (3 x 17) + ... Estos paréntesis recogen el producto del dato por la frecuencia con que aparece; incluso 11 puede verse como uno de esos productos, ya que la frecuencia de aparición del dato 11 es 1.

De modo que la suma de todos los datos, uno por uno, puede sustituirse por la

suma de los productos de cada dato por la frecuencia correspondiente. Evidentemente, ahora habrá menos sumandos; si antes eran 20 datos sueltos a sumar, ahora serán 12, de los cuales 6 serán pequeños productos (15x2, 17x3, 18x3, 19x2, 20x2, 21x2). De esta forma, la expresión adecuada para el cálculo de la media es:

$$\overline{x} = \frac{x_1.f_1 + x_2.f_2 + ... + x_k.f_k}{n} \quad [2]$$

Obsérvese que en el numerador, el subíndice del último sumando es *k* y no *n* ya que, como acabamos de decir, ahora habrá menos sumandos, porque algunos datos están repetidos (tienen una frecuencia mayor que 1). Es decir, los x_1, x_2, ..., x_k, representan ahora cada uno de los valores *distintos* de los datos.

Esta fórmula es, en realidad, una expresión muy relacionada con la [1]; de hecho, si todos los datos de una colección fueran distintos, sus frecuencias valdrían 1, *k* sería igual a *n* (¿por qué?), y la fórmula [2] se convertiría en la [1].

7. Calcule la media de los datos que aparecen en la tabla anterior, aplicando la fórmula [2]. Ahora, compare este valor con el obtenido en el ejercicio **5**. ¿Son iguales?

8. Considere ahora el segundo conjunto de datos de inasistencias a la escuela presentado anteriormente: {12, 11, 12, 13, 14, 13, 15, 16, 14, 15, 18, 16, 19, 15, 21, 20, 24, 23, 22, 23, 23, 23, 22}. Obtenga el promedio de inasistencias de la forma que usted desee.

Finalmente, nos queda el caso de hallar la media de un conjunto de ***datos*** cuando éstos se ***presentan en una tabla de frecuencias para datos agrupados en clases***. Para nuestro primer ejemplo de inasistencias:

Datos agrupados	Frecuencias (N = 20)
De 11 a 14	3
De 15 a 18	8
De 19 a 22	6
De 23 a 26	3

No podemos utilizar la fórmula [1], ya que desconocemos los datos sueltos; en principio, tampoco la fórmula [2], por la misma razón. Pero esta última fórmula sí nos brinda un camino para utilizarla en el caso de datos agrupados en clases, ya que en su cálculo intervienen frecuencias.

Pues bien, ahora también conocemos frecuencias, las de cada clase; la pregunta es: ¿por qué valor se va a multiplicar la frecuencia 3, correspondiente a la clase (11, 14) y, así, las demás frecuencias? Hay que seleccionar un valor en cada clase. ¿Qué nos dice la lógica, a falta de mayores precisiones? Que seleccionemos el valor que queda exactamente en la mitad del intervalo. ¿Cómo se calcula este valor? Sencillamente, sumando los dos extremos de la clase y dividiendo la suma entre 2. Por ejemplo, el valor intermedio en la clase (11, 14) es: (11 + 14)/2 = 25/2 = 12,5; este valor equidista de 11 y de 14.

Este nuevo valor va a representar a la clase (11, 14) a la hora de calcular la media de todos los datos; y no importa si pertenece o no al conjunto inicial de datos. A este valor intermedio de cada clase se le denomina ***marca de la clase***; vamos a representarlo por m_1, m_2, ... hasta m_j, donde *j* indica el número de clases que tenemos en la distribución de datos agrupados (en nuestro ejemplo, $j = 4$).

Con todas estas precisiones, la expresión de la media para el caso de datos agrupados en clases es:

$$\overline{x} = \frac{m_1.f_1 + m_2.f_2 + ... + m_j.f_j}{n} \quad [3]$$

Para facilitar este cálculo, podemos ayudarnos con una tabla más completa que la anterior:

Clases	Marcas de clase (m_i)	Frecuencias (f_i)	Marcas x Frecuencias (m_i x f_i)
(11, 14)	12,5	3	37,5
(15, 18)	16,5	8	132
(19, 22)	20,5	6	123
(23, 26)	24,5	3	73,5
Totales:		n = 20	366

Ahora basta con dividir entre sí los totales de las columnas 4ª y 3ª: = 366 / 20 = 18,3. Este es el promedio de inasistencias diarias durante el mes considerado. Como debe ser un número entero, podemos aproximarlo a 18.

Al comparar este valor con el obtenido en los ejercicios **5.** y **6.** (hágalo), quizá descubramos una pequeña diferencia; y es que cuando los datos están agrupados en clases, se pierde un poco de precisión (aunque se gana en otros aspectos, como ya dijimos anteriormente).

Hay otras ***formas prácticas de obtener la media*** o el promedio de un conjunto de datos. Veamos estas dos:

1. ***Utilizar la calculadora***. Si ésta posee funciones estadísticas, basta con introducir los datos y pulsar luego la tecla correspondiente a la media. Si la calculadora no posee tales funciones, podemos efectuar la suma progresiva de todos los datos (directamente o en el registro de memoria M+) y dividir el resultado final entre el número de datos.

Este es uno de los casos en que la calculadora puede servirnos como herramienta, aliviándonos del tedioso trabajo de efectuar sumas tan largas. Lo importante es conocer el significado de lo que estamos haciendo y su por qué; garantizado este conocimiento conceptual, bienvenida sea la calculadora (una vez más).

2. ***Tomar de entrada un valor imaginario para la media y luego ajustarlo con los datos***. Veamos qué significa esto. Supongamos que las edades de un grupo de 20 niños son las siguientes: 8, 7, 8, 6, 9, 7, 8, 6, 7, 7, 8, 8, 6, 10, 9, 10, 7, 8, 7 y 10 años. En este caso, tomemos 8 como valor de entrada de la media de edades de los 20 niños. Ahora recorremos ese conjunto de datos y anotamos la diferencia de cada uno de ellos con respecto a 8:

Datos	8	7	8	6	9	7	8	6	7	7
dif +					1					
dif -		1		2		1		2	1	1

8	8	6	10	9	10	7	8	7	10	Total
			2	1	2				2	8
		2				1		1		12

La suma de las diferencias positivas es 8; y la de las negativas, 12. Al compensarse entre ambas, nos queda una diferencia negativa de 4. ¿Diferencia respecto a qué? A la suma total de los 20 datos, si todos hubieran tenido el valor de la media, 8. Esta suma total hubiera sido: 20 x 8 = 160. Por consiguiente, la suma total verdadera de los 20 datos es: 160 – 4 = 156. Ahora se divide entre 20 y obtenemos la media: 7,8 años.

También podemos proceder dividiendo esa diferencia negativa final, 4, entre 20, lo que nos da el valor de 0,2; basta restar ahora este valor de la supuesta media inicial 8, con lo que obtendremos la media verdadera: 8 – 0,2 = 7,8 años.

En principio, este procedimiento puede parecer engorroso, pero la verdad es que puede hacerse mentalmente, recorriendo los datos uno por uno y compensando sucesivamente las diferencias positivas y negativas sobre la marcha: “7 me da 1 negativo; 6 me da 2 negativos, llevo 3 negativos; 9 me da 1 positivo, llevo 2 negativos; 7 me da 1 negativo, llevo 3 negativos; etc.”.

Como puede verse, este procedimiento nos da mayor soltura en el cálculo, más que si utilizáramos los datos reales, que siempre son más “pesados” de manejar. Además, tenemos libertad para elegir el valor inicial para la media que, incluso, puede no coincidir con ninguno de los datos.

Por ejemplo, podríamos haber tomado 7 como valor inicial; o incluso, 7,5... En estos casos, ¿cómo hubieran sido los cálculos de las diferencias? ¿Cómo hubieran sido las sumas finales de las diferencias? ¿Y la media verdadera? Verifíquelo, para salir de dudas... y saque sus propias conclusiones. Y trate de justificarlas. Digamos, finalmente, que el procedimiento que acabamos de describir es ideal para estimar (dar un valor aproximado de) la media de una distribución de datos.

3.2. La mediana

La mediana es el valor que, una vez ***ordenados todos los datos, se encuentra en el “medio”***, en la mitad de la distribución. Si el número de datos es impar, coincidirá con uno de los datos; si es par, puede que no ocurra esa coincidencia: hay que promediar los dos valores que se hallen en el centro de la distribución ordenada.

a) Sea el conjunto de datos: {11, 7, 10, 9, 10, 8, 7, 10, 9}

Ordenado de menor a mayor: {7, 7, 8, 9, 9, 10, 10, 10, 11}

Valor central (en la quinta posición): 9

9 es la mediana del conjunto

La mediana coincide con uno de los datos del conjunto

b) Sea el conjunto de datos: {10, 13, 12, 19, 17, 11, 15, 14, 16, 18}

Ordenado de mayor a menor: {19, 18, 17, 16, 15, 14, 13, 12, 11, 10}

Valores centrales: 15 y 14

Mediana: (15 + 14) / 2 = 14,5

La mediana no coincide con ninguno de los datos del conjunto

c) Sea la distribución de datos:

Datos	Frecuencias (N = 20)
12	4
15	6
17	8
20	2

Los datos centrales son el 10° (15) y el 11° (17)

La mediana es: (15 + 17) / 2 = 16

La mediana no coincide con ninguno de los datos de la distribución

Cuando los datos se presentan en una tabla de frecuencias para datos agrupados en clases, es suficiente con indicar la clase mediana, es decir, la clase en la que se encuentra el dato que ocupa el lugar central de la distribución. Por ejemplo, para el caso:

Datos agrupados	Frecuencias (N = 20)
De 11 a 14	3
De 15 a 18	8
De 19 a 22	6
De 23 a 26	3

la clase mediana es la que va de 15 a 18 (ahí se encuentran los datos 10º y 11º de la distribución) [puede obtenerse con más precisión un valor para la mediana, pero esto sólo nos llenaría de fórmulas más complejas que pueden verse en cualquier tratado de Estadística].

9. Calcule la mediana de las inasistencias para los datos del ejemplo inicial: {15, 19, 18, 18, 17, 17, 11, 13, 19, 18, 20, 21, 23, 26, 24, 21, 20, 17, 15, 12}.

10. Calcule la mediana de los pesos (en Kg) de los niños del ejemplo dado anteriormente: {37,8; 35,6; 34; 31,9; 40,5; 34,2; 35,6; 38,7; 32,8; 35,4; 41,6; 39,8; 34,5; 37; 42; 36,6; 31,9; 36,5; 35,7; 36; 38; 44,1; 37,2; 36,8; 35; 33,5; 38,9; 37,5; 34; 36,5; 42,5}.

11. ¿Tiene sentido calcular la mediana de los datos cuando la variable es cualitativa? ¿Por qué?



3.3. La moda

La moda es, sencillamente, el ***valor que más se repite*** (el que está más “de moda”); por lo tanto, puede haber una o más modas en la distribución de los datos. Lo que sí es cierto –a diferencia de lo que ocurre con la media y la mediana- es que la moda siempre coincide con un dato de la distribución. Y que es la única medida que puede obtenerse cuando los datos son cualitativos; por ejemplo, si se trata de la distribución por lugares de origen, etc.

Inasistencias	Frecuencia	Fre
11	1	
12	1	
13	1	
15	2	
17	3	
18	3	
19		
20		
21		
23		
24		
26		

Moda
17 y 18

a) Sea el conjunto de datos: {7, 12, 8, 7, 10, 10, 8, 9, 12, 10}
La moda es 10 (se repite 3 veces)
La distribución es unimodal

b) Sea el conjunto de datos: {8, 7, 15, 13, 7, 10, 13, 15, 9, 11}
La moda corresponde a los valores 7, 13 y 15 (se repiten dos veces cada uno)
La distribución es trimodal

Cuando los datos se presentan en una tabla de frecuencias para datos agrupados en clases, sólo se puede indicar la ***clase modal***, es decir, la clase que presenta mayor frecuencia. Por ejemplo, para el caso:

Datos agrupados	Frecuencias (N = 20)
De 11 a 14	3
De 15 a 18	8
De 19 a 22	6
De 23 a 26	3

la clase modal es la que va de 15 a 18 (presenta la mayor frecuencia, 8).

12. Indique la moda de las inasistencias para los datos del ejemplo inicial: {15, 19, 18, 18, 17, 17, 11, 13, 19, 18, 20, 21, 23, 26, 24, 21, 20, 17, 15, 12}.

13. Indique la moda de los pesos (en Kg) de los niños del ejemplo dado anteriormente: {37,8; 35,6; 34; 31,9; 40,5; 34,2; 35,6; 38,7; 32,8; 35,4; 41,6; 39,8; 34,5; 37; 42; 36,6; 31,9; 36,5; 35,7; 36; 38; 44,1; 37,2; 36,8; 35; 33,5; 38,9; 37,5; 34; 36,5; 42,5}.

Ya tenemos los conceptos de media, mediana y moda de una distribución de datos, así como la forma de calcularlos o descubrirlos. Al hallar sus valores en los ejercicios anteriores nos habremos dado cuenta de que, habitualmente, estos valores se ubican entre los valores del centro de la distribución, cuando están ordenados los datos. De ahí viene su calificativo de ***medidas de tendencia central*** –o ***valores cen-***

trales, o ***medidas centrales***- de un conjunto de datos.

3.4. Otros aspectos matemáticos de las medidas de tendencia central

Con el fin de reforzar algunos aspectos relativos al cálculo de la media y a la determinación de la mediana y de la moda, así como para destacar las relaciones entre ellas, vamos a proponer la resolución de algunos ejercicios.

¿Pueden coincidir las tres medidas centrales en una misma distribución de datos? Si su respuesta es positiva, construya un ejemplo de tal distribución. Si es negativa, explique por qué.

a) Construya ahora, si es posible, una distribución en la que no coincida ninguna de las tres medidas.
b) Ídem, en la que coincidan la media y la mediana, pero no así la moda.
c) Ídem, en la que coincidan la moda y la mediana, pero no así la media.
d) Ídem, en la que coincidan la media y la moda, pero no así la mediana.

En una clase de 6° grado hay un grupo numeroso de alumnos muy capaces. Si las calificaciones en Matemática se dan en la escala de 1 a 20, ¿qué media de calificaciones puede esperarse? ¿Y qué mediana? ¿Y qué moda? ¿Es probable que la moda sea alta?

Si la mediana de un grupo de calificaciones de Historia es 7 (en la escala de 1 a 10) y la nota aprobatoria es 5, ¿puede decirse que el grupo, en promedio, aprobó? ¿Por qué?

Veamos ahora el siguiente cuadro de posibles casos de calificaciones (en la escala de 1 a 20):

Caso	Media	Mediana	Moda
1	12	12	12
2	15	13	18
3	11	10	08
4	12	16	11

Suponga que el curso está integrado por 20 alumnos. Para cada uno de los 4 casos construya, si es posible, una distribución de datos que se ajuste a los valores dados de las medidas de tendencia central.

14. ¿Qué le ocurre (o le puede ocurrir) a la media de un conjunto de datos no todos iguales, si se elimina de dicho conjunto: a) el dato mayor; b) el dato menor; c) un dato de valor igual a la media; d) un dato de valor igual a la mediana; e) un dato de valor igual a la moda? [Responda cada pregunta con alguno de los siguientes códigos: 1. aumenta; 2. disminuye; 3. permanece igual; 4. no se puede asegurar nada].

15. ¿Qué le ocurre a la media de una distribución de datos si el número de datos es par y:
a) todos los datos aumentan en 2 unidades?
b) todos los datos disminuyen en 3 unidades?
c) la mitad de los datos aumenta en 2 unidades y la otra mitad queda igual?
d) la mitad de los datos aumenta en 3 unidades y la otra mitad disminuye en 1?

16. ¿Qué le ocurre a la mediana en los cuatro casos anteriores?

Se ha calculado la media de un grupo de 20 calificaciones. Pero posteriormente, 7 calificaciones suben en 2 puntos, 5 quedan igual, 3 disminuyen en 2 puntos, 2 disminuyen en 3, 1 disminuye en 4, y 1 disminuye en 5 puntos. ¿Qué le ha ocurrido a la nueva media con respecto a su valor anterior?

No conocemos los datos, ni tampoco el valor de la media; tampoco nos piden el valor de la nueva media, sino su variación con respecto a la anterior. Esto significa que tenemos que centrarnos en la variación que han experimentado los datos y, particularmente, el resultado final de las variaciones de todos los datos.

Calculamos los aumentos que han experimentado los datos: 7 x 2 = 14 puntos.

Y las disminuciones que han experimentado: 3 x 2 + 2 x 3 + 1 x 4 + 1 x 5 = 21 puntos.

Variación total final: La suma de los datos ha disminuido en 7 puntos.

Por consiguiente, la media habrá disminuido en 7/20 = 0,35 puntos.

Obsérvese que ***la media depende directamente de la suma de todos los datos; y que su variación depende de la variación de esta suma.***

17. Un conjunto de 300 niños se reparte en lotes de 10 niños. En cada lote, los niños tienen exactamente las siguientes edades: 8 años (3 niños); 9 años (2 niños); 10 años (1 niño); y 11 años (4 niños). ¿Cuál es la media de las edades de los 300 niños? ¿Y la mediana? ¿Y la moda?

3.5. Las medidas de tendencia central como representativas del conjunto de datos

Los ejercicios anteriores son, sin duda, útiles para desarrollar destrezas en la aplicación de los conceptos de las medidas de tendencia central, así como de los procesos para su obtención o transformación. Pero podemos formularnos otra pregunta de mayor interés: ¿Para qué se obtienen las medidas de tendencia central?

En principio, hemos dicho que son como ***representativas de todo el conjunto de datos***, aun cuando cada una lo hace a su manera. Lo que nos interesa es tener criterio para que, a la vista de cada variable y de la distribución de sus datos, podamos decidir cuál(es) de los tres valores resulta(n) más representativo(s) en cada caso. Para ello, comenzaremos por precisar las potencialidades y las limitaciones de cada una de las tres medidas centrales.

La ***media*** es la medida de tendencia central de ***mayor uso***, puesto que:

- su cálculo es sencillo;
- nos da una idea resumida y más consistente del conjunto de datos (el promedio de sus valores);
- varía en concordancia con los datos, en el sentido de que si todos los datos aumentan o disminuyen en la misma cantidad, o se multiplican o dividen por la misma cantidad, la media queda afectada por la misma variación;
- suele decirse que es más estable que la mediana; esto significa que si extraemos diversas muestras de la misma población, las medias de estas diversas muestras se parecen más entre sí que las medianas de las mismas muestras.

Pero, por otro lado, la media:

- no nos dice cuán variables son los datos, cómo difieren unos de otros;
- corre el riesgo de dejarse influir por los valores extremos de la distribución, si hay alguno(s) de ellos muy distante(s) de los demás.

Así, por ejemplo, estos tres conjuntos de datos: {11, 12, 13, 14, 15, 16, 17}, {14, 14, 14, 14, 14, 14, 14} y {2, 2, 2, 14, 26, 26, 26} tienen la misma media (verifíquelo), pero son completamente diferentes en la estructura de sus datos; la media no puede darnos una idea de la variabilidad de los datos.

En cuanto a la segunda restricción, en el conjunto de datos: {7, 5, 3, 8, 4, 5, 3, 61}, la media es: 96 / 8 = 12, valor que está afectado por el último dato (61) ya que la media de los siete primeros datos sería 35 / 7 = 5. Como se ve, el valor 12 no representa realmente al conjunto de los ocho datos.

Por su parte, la ***mediana*** también se obtiene fácilmente y no está influida por los valores extremos, aunque tampoco nos dice nada de cómo son en realidad los datos de ambas mitades de la distribución, ni acerca de su variación con respecto a la mediana.

Así, por ejemplo, observe que estos dos conjuntos ordenados de datos: {1, 3, 3, 3, 4, 14, 21, 27, 27, 27, 27} y {12, 12, 13, 14, 14, 15, 15, 16, 16} tienen la misma mediana (14), pero difieren significativamente entre sí.

Finalmente, de la ***moda*** no hay mucho más que decir, salvo insistir en su ambigüedad: puede darse incluso el caso de que no represente a un valor del centro de la distribución, sino extremo. Por ejemplo, en el siguiente conjunto de datos: {11, 13, 16, 17, 18, 20, 20, 20}, la moda es 20, que representa al mayor de los datos.

Conocer las potencialidades y limitaciones de cada uno de los valores centrales es un prerrequisito necesario para poder responder a la pregunta del para qué se obtienen las medidas de tendencia central, o de su equivalente, cuál de los valores representa mejor al conjunto de datos. Pero el condicionante principal de la respuesta está en el ***análisis de los requerimientos***

propios de cada situación, de la variable en juego y de los datos recolectados.

Supongamos que tenemos la distribución de las edades de los alumnos del salón de clase y que calculamos los valores de sus medidas de tendencia central. ¿Qué significa la media de tales edades? ¿Y la mediana? ¿Y la moda? ¿Alguna de estas medidas es más representativa del conjunto de edades que las demás? ¿O todas ellas tienen algo peculiar que aportar?

18. Para poder proceder a su posterior dotación, acabamos de obtener la distribución de datos referentes a las tallas de zapatos, camisetas, pantalones y faldas de todos los niños y niñas de la escuela. En cada una de estas cuatro distribuciones, ¿qué sentido tiene obtener la media de las tallas? ¿Y la mediana? ¿Y la moda? ¿Alguna de estas medidas es más representativa que las demás? ¿Podemos prescindir de alguna(s) de estas medidas, tomando en cuenta el objetivo de su recolección?

Invente una situación en la que Ud. va a recabar unos datos y en la que la moda sea el valor más representativo del conjunto. Análogamente para la mediana. Y, finalmente, para la media.

Para concluir este punto, permítasenos presentar dos interpretaciones referentes al promedio, una de carácter irónico y la otra, graciosamente distorsionada:

"La Estadística es la ciencia que establece que si mi vecino tiene dos carros y yo no tengo ninguno, entonces los dos tenemos un carro" (Gilbert Chesterton, novelista inglés, 1874-1936)



4. Las medidas de dispersión

Ya hemos visto que las medidas de tendencia central nos aportan información acerca de la distribución de los datos de una variable; información importante y necesaria, pero no suficiente. Como dijimos y pudimos verlo en algunos ejemplos y ejercicios, ninguna de las medidas de tendencia central, ni siquiera las tres juntas, pueden ofrecernos el detalle de los datos, la variabilidad presente en ellos.

En efecto, consideremos estos dos conjuntos de datos ordenados, referentes a las calificaciones escolares de dos grupos de 20 alumnos (en una escala de 1 a 20 puntos):

Conjunto 1: {3, 4, 4, 4, 5, 5, 6, 6, 7, 10, 10, 10, 10, 15, 16, 16, 17, 17, 17, 18}
Conjunto 2: {9, 9, 9, 9, 9, 9, 10, 10, 10, 10, 10, 10, 10, 10, 10, 10, 12, 12, 12}

La media, la mediana y la moda de ambos conjuntos es la misma (10 puntos); y sin embargo, las distribuciones son muy diferentes. El segundo grupo de calificaciones se presenta más homogéneo, las notas están más agrupadas; cosa que no ocurre en el primer conjunto, en el que la dispersión de las notas es muy marcada.

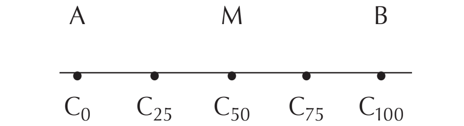
De modo que no basta con que intentemos caracterizar o resumir una distribución de datos aportando sus medidas de tendencia central; necesitamos también decir algo

con respecto a la variabilidad, a la dispersión de sus datos. Los indicadores de esta variabilidad reciben el nombre de ***medidas de dispersión***. Veamos algunas de ellas.

Lo primero que tendemos a observar son los ***valores extremos***; con ellos calculamos su diferencia que, como ya sabemos, se denomina ***rango***. El rango de una distribución de datos es la primera medida de dispersión y la más básica. En nuestros ejemplos, el rango del primer conjunto es 15 y el del segundo, 3. Nótese que si agregamos este dato al de los valores centrales, ganamos en comprensión acerca de cada uno de los dos conjuntos de datos y podemos diferenciarlos de inmediato.

Pero si bien el segundo conjunto queda casi fotografiado, no ocurre así con el primero, ya que su rango es grande. No sabemos cómo son los datos que ocupan lugares cercanos a los valores extremos, si son parecidos a ellos o no. ¿Cómo se puede resolver esto? Una de las maneras sencillas consiste en abrir unas "ventanitas" para ver los datos de tanto en tanto.

Por ejemplo, podemos imaginar todos los datos ordenados y colocados a la misma distancia unos de otros sobre un segmento, que va desde el punto A (el dato menor) hasta B (el dato mayor). En el punto medio de AB debe aparecer el punto M (la mediana). Pues bien, si marcamos también los puntos medios de AM y MB, tendremos señalados cinco puntos (cinco ventanitas) de observación sobre el conjunto de datos, desde C_0 hasta C_{100}:



¿Por qué se escriben los subíndices 25, 50, 75 y 100? Veamos; estos cinco puntos han dividido el segmento en cuatro partes de igual medida. Si desde C_0 hasta C_{100} se halla incluido el 100% de los datos, en cada uno de los cuatro tramos estará agrupado el 25% de los mismos. Así:

- desde C_0 hasta C_{25} está el 25% de los datos, desde el inicio;
- desde C_0 hasta C_{50} está el 50% de los datos, desde el inicio;
- desde C_0 hasta C_{75} está el 75% de los datos, desde el inicio;
- desde C_0 hasta C_{100} está el 100% de los datos.

Los valores de la distribución que ocupan esos puntos particulares reciben el nombre de cuartiles (porque marcan la división de la distribución en cuatro partes congruentes). En particular, la mediana coincide con el cuartil C_{50}.

En nuestros dos conjuntos, que constan de 20 datos, cada uno de los cuatro tramos entre los cuartiles (se denominan ***recorridos intercuartílicos***) tendrá 5 datos. Por consiguiente, los cuartiles C_{25} y C_{75} están ubicados entre las posiciones 5^a y 6^a de la distribución y entre las posiciones 15^a y 16^a, respectivamente; por lo tanto, hay que obtener el promedio de los valores que se hallan en esas posiciones.

Así, en el primer conjunto, $C_{25} = (5 + 5)/2 = 5$; y $C_{75} = (16 + 16)/2 = 16$. Y en el segundo conjunto, $C_{25} = (9 + 9)/2 = 9$; y $C_{75} = (10 + 10)/2 = 10$. De más está decir que los cuartiles pueden no coincidir con los datos de la distribución (ya lo sabíamos de la mediana). Obsérvese en particular que entre los cuartiles C_{25} y C_{75} se halla ubicado el 50% de los datos centrales de cualquier distribución.

Debemos recalcar la utilidad de estos nuevos datos; ahora sabemos que si los dos conjuntos ordenados de datos se fragmentan en cuatro partes, los valores que vamos a encontrar son: 3, 5, 10, 16 y 18 en el primer conjunto, y 9, 9, 10, 10 y 12 en el segundo, respectivamente. Aunque todavía algo borrosa, tenemos una mejor "fotografía" de cada uno de los dos conjuntos, particularmente del primero...

Otra medida de dispersión de características similares a la de los cuartiles y que se aplica cuando la distribución consta de muchos datos, es la de los ***percentiles***. En este caso, se divide el segmento, no en cuatro, sino en cien partes congruentes y se procede de una manera similar a la de los cuartiles.

Así, el percentil 80 (P_{80}) representa el valor que corresponde al 80% de todos los datos ordenados. Para hallarlo en nuestro segundo conjunto, planteamos la siguiente regla de tres:

dato en la posición nº	% correspondiente
20	100
x	80

de donde: $x = 80 \times 20 / 100 = 16$. Es decir, en nuestro caso, el 80% de los datos está contenido en el lote que va desde el primer dato hasta el dato que ocupa la posición 16. Para hallar su valor hay que obtener el promedio de los valores que se hallan en las posiciones 16ª y 17ª que son 10 y 12; de modo que: $P_{80} = (10 + 12)/2 = 11$.

19. Halle los percentiles P_{20}, P_{40}, P_{60} y P_{80} de los datos del primer conjunto

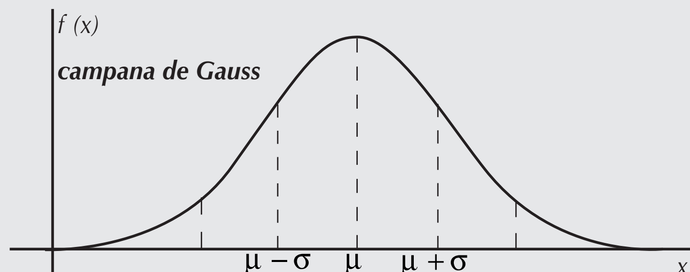
En este punto ya debe quedar claro que podemos analizar un conjunto de datos referentes a una variable, bien sea a partir de sus tablas de distribución de frecuencias, o de sus gráficas; también podemos servirnos de sus medidas de tendencia central, pero éstas deben ir acompañadas de sus medidas de dispersión.

Algunas distribuciones de datos tienen características particulares que son compartidas por muchas variables de naturaleza muy diversa (fisiológica, económica, social, psicológica...).

Entre ellas destaca la ***distribución normal***, así llamada porque los datos se agrupan simétricamente a ambos lados de la media. Algunas de las muchas variables cuyas distribuciones de datos en poblaciones numerosas siguen esa forma de distribución, son la estatura y el peso de individuos adultos, los efectos producidos por un fármaco en enfermos, o por un abono en las plantas, etc.

Su representación gráfica tiene forma de campana (***campana de Gauss***). En la gráfica, los valores de la variable *x* se colocan en el eje horizontal; μ (mu: letra griega equivalente a nuestra m) designa la media; σ (sigma: letra griega equivalente a nuestra s) designa la medida de dispersión conocida como desviación típica.

En este tipo de distribución, el intervalo de valores de x que va desde $\mu - \sigma$ hasta $\mu + \sigma$ (algo así como antes desde C_{25} hasta C_{75}...) encierra el 68% de los datos centrales de la distribución.



No vamos a continuar con este estudio; pero sí tomamos nota de que existen otras medidas de dispersión y, sobre todo, ***tipos de distribuciones***, cuyo estudio constituye el objetivo de una Estadística más avanzada.

5. Finalmente, unos ejercicios de interpretación

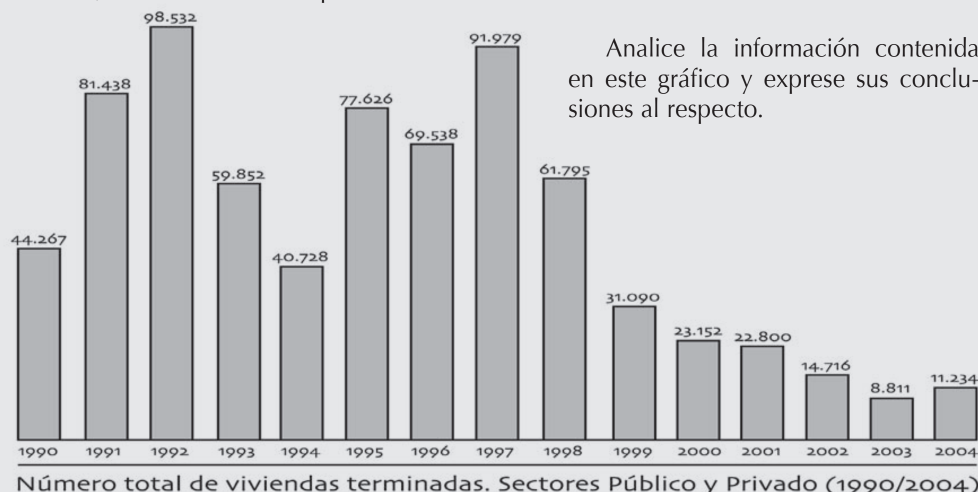
Vamos a mostrar algunos gráficos aparecidos en diversas publicaciones periódicas, con el fin de que nos ejercitemos en su interpretación.



La gráfica se refiere a la relación *"salario mínimo / salario promedio"* en varios países latinoamericanos; para calcular esa relación se hace la división correspondiente. Por ejemplo, si el salario mínimo en un país es de 4.500 pesos y el salario poblacional promedio es de 7.200 pesos, la relación vale: 4.500/7.200 = 0,625.

Obsérvese que cuanto más cerca de 1 está ese cociente, significa que el salario promedio y el salario mínimo son casi iguales. Esto puede significar dos cosas: que el salario mínimo es muy alto (cosa poco probable en nuestra región), o que el país es muy pobre, por cuanto la mayoría de los trabajadores del sector formal de la economía sólo recibe el ingreso más bajo de la escala de sueldos. Revise ahora los datos de la tabla, analice la información contenida y exprese sus conclusiones.

La siguiente gráfica corresponde al número total de *viviendas terminadas* en el período indicado, en uno de nuestros países:

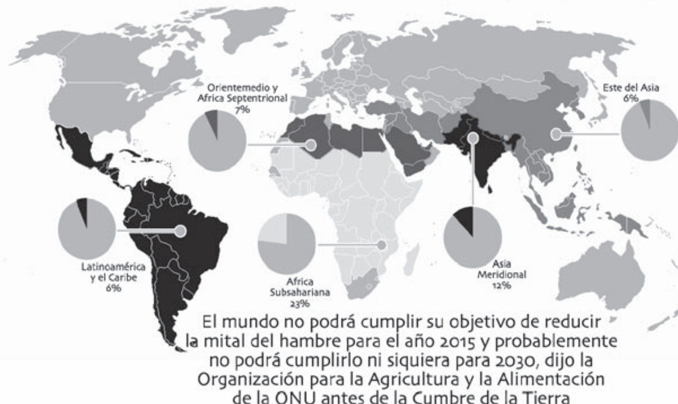


Analice la información contenida en este gráfico y exprese sus conclusiones al respecto.

Esta representación gráfica se refiere al problema del *hambre en el mundo*:

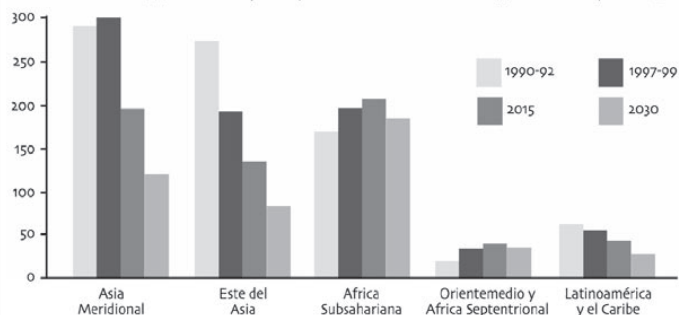
HAMBRE EN EL MUNDO

Porcentaje de habitantes desnutridos por región para 2015



Habitantes desnutridos por región entre 1990-92 y 2030

El número de personas con hambre en los paises en vías de desarrollo probablemente descienda de los 777 millones que hay en la actualidad a unos 440 millones para 2030



Fuente FAO

REUTERS

Analice la información contenida en esta representación y exprese sus conclusiones al respecto.

Veamos este otro cuadro, en el que se presentan las calificaciones merecidas por nuestros países en cuanto a las *políticas públicas desde 1980*:

Calificaciones en cuanto a las Políticas Públicas desde 1.980

País	Estabilidad	Adaptabilidad	Ejecución y cumplimiento	Coordinación y coherencia	Orientación al interés público	Eficiencia	Indice general de políticas
Argentina	Bajo	Medio	Bajo	Bajo	Medio	Bajo	Bajo
Bolivia	Medio	Alto	Medio	Medio	Medio	Medio	Medio
Brasil	Alto	Alto	Alto	Alto	Medio	Medio	Alto
Chile	Alto	Alto	Alto	Alto	Alto	Alto	Muy Alto
Colombia	Alto	Alto	Alto	Medio	Medio	Medio	Alto
Costa Rica	Alto	Medio	Alto	Medio	Alto	Alto	Alto
Rep. Dominicana	Medio	Medio	Medio	Medio	Bajo	Medio	Medio
Ecuador	Bajo	Medio	Bajo	Bajo	Bajo	Bajo	Bajo
El Salvador	Alto	Alto	Alto	Medio	Medio	Alto	Alto
Guatemala	Medio	Medio	Bajo	Medio	Bajo	Medio	Bajo
Honduras	Alto	Medio	Medio	Medio	Bajo	Medio	Medio
México	Alto	Medio	Alto	Medio	Medio	Alto	Alto
Nicaragua	Medio	Medio	Medio	Bajo	Bajo	Medio	Bajo
Panamá	Medio	Bajo	Medio	Bajo	Bajo	Bajo	Bajo
Paraguay	Medio	Bajo	Bajo	Bajo	Bajo	Bajo	Bajo
Perú	Medio	Medio	Medio	Medio	Medio	Medio	Medio
Uruguay	Alto	Alto	Alto	Medio	Medio	Medio	Medio
Venezuela	Bajo	Bajo	Medio	Bajo	Medio	Bajo	Bajo

Fuente/Informe del Banco Interamericano de Desarrollo, noviembre de 2005

EL NACIONAL

Analice la evaluación que se hace de su país y compárela con las de países vecinos al suyo. Exprese sus conclusiones al respecto.

6. *Y ahora, otros ejercicios "para la casa"...*

20. Al comenzar el año escolar, usted recolecta los datos de la edad de sus alumnos y halla los correspondientes valores centrales (media, mediana y moda). Si al comenzar el siguiente curso escolar no se ha modificado la nómina de estos alumnos, ¿qué habrá ocurrido con la media, la mediana y la moda de estos nuevos datos, con respecto a los del curso anterior?

21. Se ha calculado la media de un grupo de calificaciones. Posteriormente, la mitad de las calificaciones aumenta en 2 puntos cada una, 4 disminuyen en 1 punto, 8 disminuyen en 3 puntos, y el resto quedan iguales. Si la media del grupo no varía con respecto a la obtenida antes de estos cambios, ¿de cuántas calificaciones estamos hablando?

22. Un grupo de 300 niños se distribuye en lotes de 10 niños del siguiente modo: 8 lotes de niños de 7 años; 2, de 9 años; 7, de 8 años; 5, de 10 años; 5, de 12 años; y 3, de 11 años. ¿Cuál es la media de las edades de los 300 niños? ¿Y la mediana? ¿Y la moda?

Tome el conjunto de datos correspondientes a las edades de los alumnos de su salón y obtenga los valores de las tres medidas de tendencia central. Analice estos valores y su significado. Compárelos con los de otros salones similares (del mismo grado). Establezca sus conclusiones.

Una de las actividades más interesantes para aplicar los conocimientos estadísticos y para fomentar el espíritu indagador propio y de los alumnos, es la elaboración y aplicación de ***encuestas*** entre los mismos niños y en su entorno familiar y comunitario. A este respecto, seleccione algunos temas que considere de interés, elabore una encuesta, aplíquela, organice la presentación de los datos recolectados, y analice sus resultados.

Una de las fuentes más destacadas de informaciones en formato estadístico es la prensa periódica (diarios, revistas...). Tome los periódicos más recientes, busque algunos informes de esa naturaleza y analícelos.

¿Cuál es el color más usado en las banderas (sin considerar los escudos que algunas incluyen) de los países de Latinoamérica? Este dato, ¿representa una media, una mediana o una moda en la distribución de los colores de las banderas?

Trate ahora de resolver estos dos problemas, ya presentados en el Cuaderno no 11 (y resueltos en sus páginas 22 y 23), relativo a razones y proporciones:

ñ) Un grupo de hombres y de mujeres declaran su edad por escrito, y se calculan los promedios de esas edades: el del grupo total, es de 40 años; el de los hombres, 50 años; y el de las mujeres, 35 años. ¿Cuál es la razón del número de mujeres al número de hombres?

o) Una persona desea darse un baño con agua a 35º C. Para conseguir esa temperatura, debe mezclar agua caliente con agua fría en una determinada proporción. Hace dos pruebas: en la primera, mezcla 1 parte de agua caliente con 2 de agua fría, y obtiene agua a 20º C; en la segunda mezcla 3 partes de agua caliente con 2 de agua fría, y obtiene agua a 28º C. Con estos datos, ¿en qué proporción debe mezclar ambos tipos de agua?

Referencias bibliográficas y electrónicas

• Batanero, C. (2002). *Los retos de la cultura estadística*. Jornadas Interamericanas de Enseñanza de la Estadística. Buenos Aires. Disponible en: http://www.ugr.es/~batanero/publicaciones.htm

• CEPAL - ECLAC (2005). *Anuario Estadístico de América latina y el Caribe*. Disponible en: http://www.eclac.cl/badestat/anuario_2004/esp.htm

• Moore, D. (1998). Incertidumbre. En L. Steen (Ed.), *La enseñanza agradable de las matemáticas*, pp. 103-148. México: Limusa.

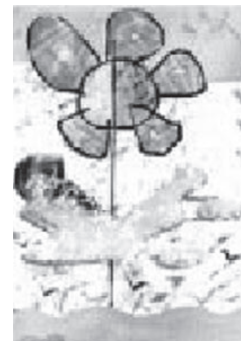
• Shadian, R. (1998), *Lectura e interpretación de gráficas de datos*. Disponible en: http://msip.lce.org/~quiz/quizzes/jahumada/u4s2a.html

Respuestas de los ejercicios propuestos

1. Cualitativas: a, c, f, h, m; cuantitativas discretas: e, k, l; cuantitativas continuas: b, d, g, i, j **2.** No. Porque las variables cualitativas no se miden en escalas de orden **3.** a) 7.600 pesos; b) 1.500 pesos **4.** No. Porque las variables cualitativas no se miden en escalas de orden **5.** 18,2 inasistencias **6.** 36,84 kg **7.** 18,2 inasistencias; son iguales **8.** 18,6 inasistencias **9.** 18 inasistencias **10.** 36,5 kg **11.** No. Porque las variables cualitativas no se miden en escalas de orden **12.** Hay dos modas (bimodal): 17 y 18 inasistencias **13.** Hay cuatro modas (polimodal): 31,9; 34; 35,6 y 36,5 kg **14.** a) 2; b) 1; c) 3; d) 4; e) 4 **15.** a) aumenta en 1 unidad; b) disminuye en 3 unidades; c) aumenta en 1 unidad; d) aumenta en 1 unidad **16.** Lo mismo que a la media en los casos a) y b); en los casos c) y d) no se puede asegurar nada **17.** Media: 9,6 años; mediana: 9,5 años; moda: 11 años **18.** No tiene sentido hallar la media y la mediana de las tallas; la medida de tendencia central que más puede interesar es la moda **19.** P_{20} = 4,5 puntos; P_{40} = 6,5 puntos; P_{60} = 10 puntos; P_{80} = 16,5 puntos **20.** Han aumentado las tres medidas en 1 unidad **21**. 28 calificaciones **22.** Media: 9,1 años; mediana: 8,5 años; moda: 7 años

Índice

A modo de Introducción	**5**
1. El significado de la Estadística	**6**
2. ¿Qué hacemos con los datos?	**8**
3. Las medidas de tendencia central	**16**
4. Las medidas de dispersión	**24**
5. Finalmente, unos ejercicios de interpretación	**26**
6. Y ahora, otros ejercicios "para la casa"...	**29**
Referencias electrónicas	**30**
Respuestas de los ejercicios propuestos	**31**




FE Y ALEGRIA